Supplemental Material

The information described below provides greater detail of the methodology and results described in the manuscript.

1. Extracting and estimating the home locations of Twitter users

The use of 'geo-tags' on Twitter—where geographical coordinates are embedded in the metadata—is relatively rare and unevenly distributed, which makes geo-tags an unreliable way to determine the locations of users.[1] A common alternative is to use the locations that users self-report in their profiles (free text).[2-4] Among the set of tweets collected for this study, 0.5% (1,735 of 358,194 tweets) included geo-tag information, while 70.1% (90,658 of 129,286 users) had some self-reported information about location in their user profiles.

Location inference methods were used to identify users located in Australia, Canada and United Kingdom. Nominatim,[5] a gazetteer, was used to translate the locations of Twitter users to identify users in the three countries. This information was taken from the tweet metadata (geo-tags) or user profile information (free text). Pre-processing steps for the user profile information included the removal of punctuation, numeric values, characters for non-English languages, and one-character words. Nominatim produces a score for the set of possible locations it returns, and a score of 0.4 was used as a threshold to avoid locations likely to be spurious (this threshold was determined through experiments in previous work). Where users included geo-tag information in the tweets they posted, the most frequent location was chosen (or the earliest where there were equally frequent and different locations used). Where users did not include geo-tag information, they were assigned to the location produced by Nominatim based on their user profile information.

2. Construction of the follower network

The social connections among the set of 16,789 users who posted about HPV vaccines and were located within Australia, Canada, and the UK were examined. The follower connections to and from each of the 16,789 users were collected through calls to the Twitter Application Program Interface (API), performed shortly after the first time each users posted a relevant tweet during the relevant time period. These data were used to construct the internal follower network by reconciling connections to and from each user to any other user in the set.

This study evaluated the proportions of international connections across the three countries, and examined the differences in the proportions of users who mostly post tweets expressing concerns about HPV vaccines relative to all other users. To do this, the ratios of follower connections of two types of users (those who express concerns in at least half of their relevant tweets versus all other users) in the three countries (Australia, Canada, and the UK), were compared.

3. Machine learning methods used to train and test the classification of the tweets

3.1. Pre-processing

The tweet texts were processed to construct features for our classifiers. No modifications were made on words that were hashtags (beginning with “#”) and Twitter usernames (beginning with “@”). If a tweet text contained a website link (URLs), the domain name was stored. Standard data pre-processing including the removal of common English words,[6,7] and the removal of plurals and modifiers using Porter algorithm.[8] All numerical values were removed and all words were converted into lowercase. Each tweet was then transformed into a binary representation—a vector of length equal to the total number of unique features found across all tweets—with 1 marked for any feature in the tweet and 0 for all other features.

3.2. Supervised Machine Learning Training

Due to the large number of tweets collected in the period, a supervised machine learning approach was used to classify the tweets. This involved the manual labeling of a random sample of tweets, which were then used to train algorithms to identify similar patterns in the remaining tweets.[9,10] The data were classified in two stages to firstly identify tweets expressing concerns, and then to classify those concerns by type (Figure A1).

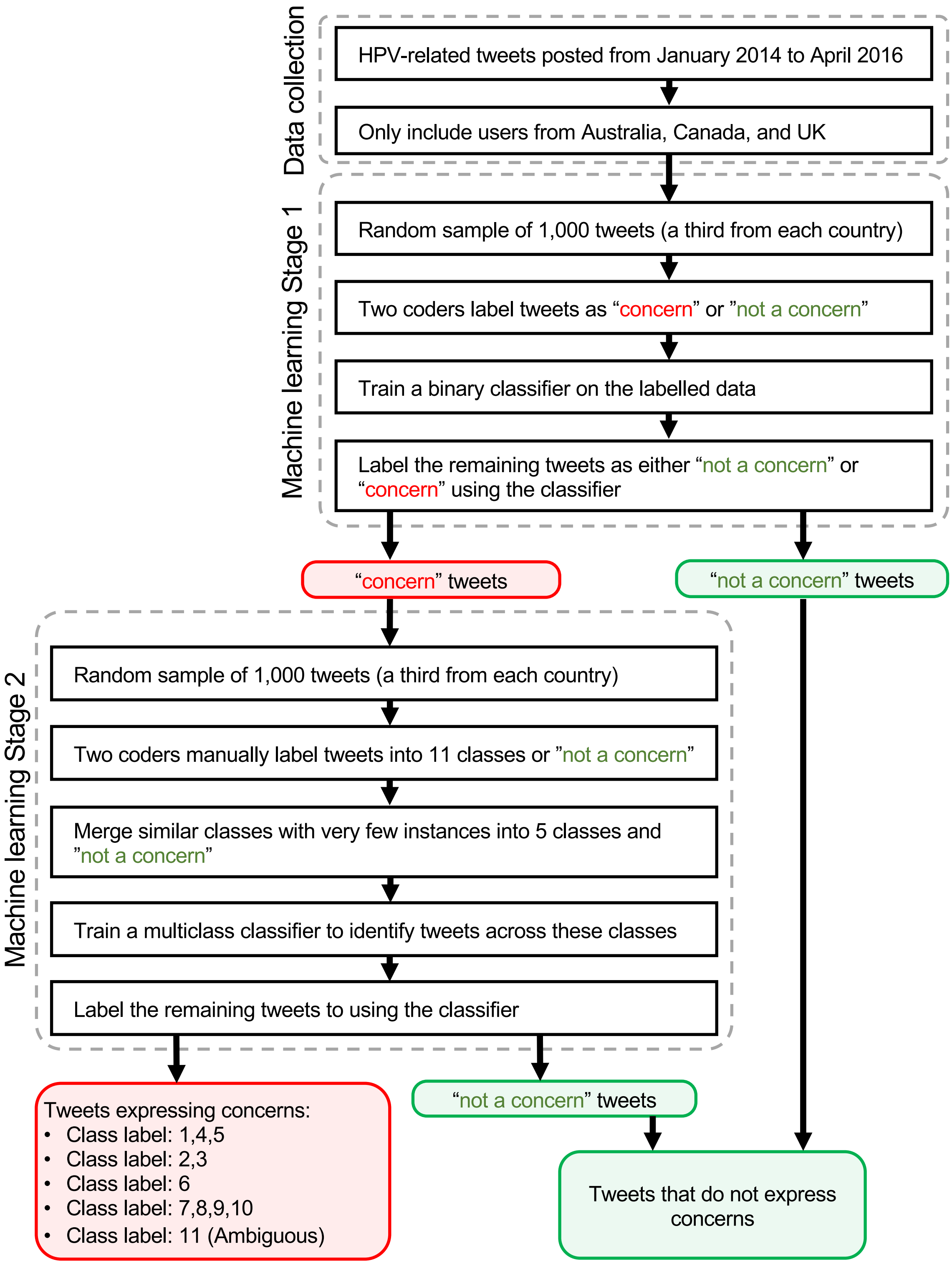

Figure A1. The design of the two-stage method for identifying classes of concerns in HPV vaccine tweets, where Stage 1 is the construction of a binary classifier and Stage 2 is the construction of a multiclass classifier

Classifiers for identifying any concern

The first stage of coding aimed to distinguish between tweets about HPV vaccines that expressed a concern versus tweets not expressing a concern. A random sample of 1,000 tweets were manually labeled as ‘concern’ or ‘non-concern’ by two investigators to form a training set from which to train a machine learning classifier. A separate set of 150 tweets was presented to the coders as a practice set for discussion prior to the independent labeling process. If the tone of the tweet was not immediately clear from the information provided within the tweet, the coders used links to webpages if they were available. There was strong agreement amongst the coders (92.3% agreement, Cohen’s κ=0.81; 95% CI 0.77-0.85), and disagreements were resolved by discussion to produce the final training set.

A supervised binary classifier was trained using the manually labeled tweets to assign labels to the rest of the tweets in the data set. This study used a linear support vector machine (SVM) with stochastic gradient descent learning method to perform binary classification. The SVM method has been used widely for applications that deal with unbalanced and high dimensional data sets like those described here.[11-15] In this study, a random sample of 80% of the manually labeled tweets (the *training set*) was used to train and validate the classifier and the remaining 20% was used to test the performance of the classifiers (the *testing set*). The best parameters for the classifier were chosen using 10-fold cross validation using the training set. K-fold cross validation is a common method used to train a classifier in a prediction problem, where the training set is partitioned into K equal sized subsamples.[16] During the training process, the classifier is trained using all but one of the subsamples and validated on the remaining subsample, repeating the process K times. To avoid overfitting, the L2 regularization was used with 1,000 iterations during the training.[17,18]

Classifiers for identifying specific concerns

The second stage of coding aimed to distinguish different types of concern. The tweets classified as having expressed a concern about HPV vaccines in stage one were examined to distinguish the specific types of concerns. A random sample of 1,000 tweets was selected and a separate set of 150 tweets was used to pilot and test the scheme prior to the independent labeling process.

The categories for types of tweets expressing concerns were determined using an inductive and deductive procedure. Accordingly, the Health Belief Model (HBM) and additional concerns towards the HPV vaccine that have been identified in the literature were used to develop an initial coding scheme of 12 types of concerns (Table A1). These coding categories were discussed and agreed upon amongst the research team. There was good agreement in coding the random sample of 1,000 tweets (79.0% agreement, Cohen’s κ=0.71; 95% CI 0.67-0.74), and any disagreements were resolved by discussion to produce the final training set. For example, the HBM factor of ‘self-efficacy’ was originally included in this coding scheme but deleted after team consultation as it overlapped with other groups during rounds of practice coding (i.e. ‘logistical barriers’).

Table A1. Number of coded tweets of each type of concern

Class label	Description	Number of tweets
1	Not beneficial	8
2	Perceived logistical challenges	18
3	Perceived harms	462
4	Not severe	3
5	Low susceptibility	1
6	Cues to action	141
7	Mistrust	90
8	Undermining religious principles	11
9	Undermining civil liberties	58
10	Additional concerns not otherwise specified	6
11	Tweet is ambiguous	18
12	No concern expressed	184
Total		1000

As can be seen in Table A1, some of the classes had fewer than 10 examples identified. Rare classes of concerns were merged based on similar themes to provide enough relevant examples to train and evaluate the performance of the multi-class classifier (Table A2).

Table A2. Number of coded tweets of each type of concern after merging the labels

Class label	Original class labels	Number of tweets
Unnecessary	1,4,5	12
Perceived barriers	2,3	480
Cues to action	6	141
Additional concerns	7,8,9,10	165
Ambiguous	11	18
Non-concern	12	184

In the second stage, the machine learning task was a multi-class classification. A *one-versus-rest* strategy was adopted where tweets from one class (type of concern) were treated as positive samples and all other tweets were treated as negative samples. A single linear support vector machine (SVM) with stochastic gradient descent learning method was trained as the classifier for each class and this was repeated for all types of concerns. The final label for each tweet was assigned to the class for which there was the highest likelihood of it belonging to the positive class. Given the unbalanced nature of the labeled data (some classes have a large number of tweets while several others have a small number of tweets), a random sample of 65% of the labeled tweets (the *training set*) were used to train the classifiers and the remaining 35% of the labeled tweets (the *testing/holdout set*) were used to test the performance of the classifiers. The class weights were adjusted to be inversely proportional to the number of tweets in the classes in order to mitigate the influence effect of large classes during the training. The best parameters for the classifiers were chosen using the same approach as described above.

4. Proportional exposure to HPV vaccine related tweets

The total number of followers each of the users had at the time they posted their tweets was also used to measure the potential exposure to those tweets and the potential size of the audience for each class of concerns expressed by users within each country. To quantify the potential exposure to tweets by country and type of concern, the potential exposure to each tweet was defined by the number of followers that a user had at the time they posted a tweet about HPV vaccines.

Tweets expressing concerns tended to have smaller audiences compared with tweets not expressing concern about HPV vaccines (Tables A3 and A4). In Canada, tweets expressing concerns had a total potential exposure count of 3.75% (4.81 million of 128.4 million total potential exposures to tweets from users in Canada). In Australia, the proportion was 11.0% (3.25 million of 29.7 million total potential exposures to tweets from users in Australia), and in the UK, the proportion was 16.3% (21.3 million of 130.4 million total potential exposures to tweets from users in the UK).

The difference between the number of tweets and the relative sizes of the audiences show that expressions of concern about HPV vaccines were likely to have reached a smaller overall audience than would be expected given the number of tweets. Note that these numbers reflect the total number of exposures to each type of tweet rather than the total number of unique users who may have seen those tweets.

Table A3. The number and proportion of exposures to tweets classified as expressing concerns, by country

Country	Concern	Non-concern	Total
Australia (%)	3,254,528 (10.97%)	26,422,799 (89.03%)	29,677,327 (100%)
Canada (%)	4,810,618 (3.75%)	123,608,306 (96.25%)	128,418,924 (100%)
UK (%)	21,260,539 (16.30%)	109,182,707 (83.70%)	130,443,246 (100%)
Total	29,325685 (10.16%)	259,213,812 (89.84%)	288,539,497 (100%)

Table A4. The number and proportion of exposures to tweets posted by users, by country and type of concern

Group label	Australia (%)	Canada (%)	UK (%)	Total (%)
Unnecessary	7,508 (0.03%)	9,511 (0.01%)	53,384 (0.04%)	70,403 (0.02%)

Perceived barriers	1,950,348 (6.57%)	2,140,893 (1.67%)	9,912,306 (7.60%)	14,003,547(4.85%)
Cues to Action	287,686 (0.97%)	390,540 (0.30%)	803,537 (0.62%)	1,481,763 (0.51%)
Other concerns	595,832 (2.01%)	1,086,688 (0.85%)	1,978,772 (1.52%)	3,661,292 (1.27%)
Ambiguous	413,154 (1.39%)	1,182,986 (0.92%)	8,512,540 (6.53%)	10,108,680 (3.50%)
All non-concern	26,422,799 (89.03%)	123,608,306 (96.25%)	109,182,707 (83.70%)	259,213,812 (89.84%)
Total	29,677,327 (100%)	128,418,924(100%)	130,443,246 (100%)	288,539,497 (100%)

5. Performance of the classifiers

The binary classifier was designed to distinguish between tweets about HPV that expressed concerns from non-concerns. The binary classifier produced a precision of 90% and a recall of 90% (Table A5). In other words, approximately 1 in 10 tweets expressing a concern could have been misclassified as a non-concern tweet, and approximately 1 in 10 tweets not expressing a concern could have been misclassified as a tweet expressing a concern. Analyses reported should be interpreted in the context of this accuracy.

Table A5. Performance measures for the binary classifier within the testing/holdout set

Class label	Precision	Recall	F1 score	Number of tweets in the test set
Concern	0.90	0.97	0.93	143
Non-concern	0.89	0.74	0.81	57
Average/Total	0.90	0.90	0.90	200

The performance of the multi-class classifier varied relative to the number of instances available for training and testing in the labeled set of 1000 tweets (Table A6). The precision and recall were over 90% when identifying tweets from the ‘cues to action’ group, but a substantial proportion of tweets from other classes were misclassified as Class 11 (ambiguous tweets) when testing the classifier on the holdout. The performance results suggested that one could be reasonably confident about the proportions of tweets in ‘perceived barriers’ and ‘cues to action’ groups, but less confident about the proportions of tweets belonging to other classes.

Table A6. Performance measures for the multi-class classifier within the testing/holdout set

Class label	Precision	Recall	F1 score	Number of tweets in the test set
Not beneficial, not severe, & low susceptibility	0.50	0.14	0.22	7
Perceived barriers	0.81	0.74	0.77	180
Cues to Action	0.91	0.92	0.92	53
Other concerns	0.77	0.46	0.57	50

Ambiguous	0.03	0.4	0.06	5
Non-concern	0.45	0.31	0.37	55
Average/Total	0.74	0.64	0.68	350

6. Examination of the follower network

Examining the followers of users who expressed concerns about HPV vaccines, the results show that 34.7% of the followers of users expressing concerns were also sharing their concerns. In contrast, 8.3% of the followers of users who did not express concerns were users expressing concerns (Table A7).

Table A7. Aggregate percentages of followers for all countries and expression of concern

	Concern (%)				Non-concern (%)			
Internal network followers (aggregate follower count)	Australia	Canada	UK	All	Australia	Canada	UK	All
All concern (38,378)	4.5	17.8	12.4	34.7	9.9	18.4	37.0	65.3
All non-concern (464,251)	1.5	2.3	4.5	8.3	20.4	25.2	46.1	91.7
All users (502,629)	1.7	3.5	5.1	10.3	19.6	24.6	45.4	89.7

Examining the followers of users who expressed concerns about HPV vaccines, the results also show that these users were relatively well connected to users in other countries who also expressed concerns (Table A8). For example, 28.6% of the followers of Australian users expressing concerns were also users expressing concerns, and 52.4% of those followers were from Canada or the UK.

This type of social connection—between users from different countries—was disproportionately high between users expressing concerns about HPV vaccines, and this pattern was consistent across the three countries. These differences are also apparent in Figure 1 in the manuscript, where there is a higher density of users expressing concerns about HPV vaccines close to the boundaries between the clusters of users from Canada and the UK.

Table A8. Aggregate number and percentage of followers by country and expression of concern

	Concern (%)					Non-concern (%)				
Internal network followers (aggregate follower count)	Australia	Canada	UK	All	Proportion of international followers in the same concern group	Australia	Canada	UK	All	Proportion of international followers in the same concern group
Australian (102,894)	5.7	1.1	0.8	7.6		82.3	6.0	4.1	92.4	
-Concern (5,319)	13.6	8.3	6.7	28.6	52.4	54.4	6.5	10.5	74.4	23.8

-Non-concern (97,575)	5.3	0.7	0.4	6.4	17.2	83.8	6.0	3.7	93.6	10.4
Canadian (151,179)	0.9	9.6	1.6	12.0		6.6	71.8	9.6	88.0	
-Concern (14,656)	3.9	37.4	6.3	47.6	21.4	4.0	41.0	7.5	52.4	21.9
-Non-concern (136,523)	0.5	6.6	1.1	8.2	19.5	6.8	75.1	9.9	91.8	18.2
UK (248,556)	0.5	0.9	9.0	10.4		1.6	3.7	84.3	89.6	
-Concern (18,403)	2.3	5.0	18.8	26.1	28.0	1.8	3.9	63.1	73.9	8.3
-Non-concern (230,153)	0.4	0.5	8.2	9.1	9.9	1.6	3.7	85.6	90.9	5.8

References

1. Sloan L, Morgan J. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PloS one*. 2015;10(11):e0142209.
2. Elmongui HG, Morsy H, Mansour R. Inference models for twitter user's home location prediction. IEEE/ACS International Conference of Computer Systems and Applications (AICCSA); 2015.
3. Bo H, Paul C, Timothy B. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research*. 2014;49:451-500.
4. Bo H, Paul C, Timothy B. A. Stacking-based approach to twitter user geolocation prediction. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistic; 2013.
5. Nominatim: OpenStreetMap Wiki. http://wiki.openstreetmap.org/wiki/Nominatim.
6. Lehmann J, Gonçalves B, Ramasco JJ, Cattuto C. *Dynamical Classes of Collective Attention in Twitter*. 2012.
7. Zhao WX, Jiang J, Weng J, et al. *Comparing Twitter and Traditional Media using Topic Models*. 2011.
8. Porter MF. *An algorithm for suffix stripping*. 1997.
9. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *Journal of medical Internet research*. 2015;17(6):e144.
10. Surian D, Nguyen QD, Kennedy G, Johnson M, Coiera E, Dunn GA. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *Journal of medical Internet research*. 2016;18(8):e232.
11. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Studies in health technology and informatics*. 2015;216:761-765.
12. Mullen T, Collier N. *Sentiment analysis using support vector machines with diverse information sources*. Barcelona Spain,2004.
13. Pang B, Lee L, Vaithyanathan S. *Thumbs up? Sentiment classification using machine learning techniques*. Philadelphia United States,2002.

14. Sebastiani F. Machine learning in automated text categorization. *ACM Comput Surv*. 2002;34:1-47.
15. Joachims T. *Transductive inference for text classification using support vector machines*. Bled Slovenia,1999.
16. McLachlan GJ, Do K, Ambroise C. *Analyzing Microarray Gene Expression Data*. Wiley; 2004.
17. Coates A, Lee H, Ng A. An analysis of single-layer networks in unsupervised feature learning. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTAT); 2011.
18. Koshiba Y, Abe S. Comparison of L1 and L2 Support Vector Machines. Proceedings of the International Join Conference on Neural Networks; 2003.